Reducing Incidence of Nonpositive Definite Covariance Matrices in Mixed Effect Models

Daniel McNeish[1] & Daniel J. Bauer[2]

[1]Arizona State University, USA
[2] University of North Carolina, Chapel Hill, USA

Published in *Multivariate Behavioral Research* on October 14, 2020

This work was supported by Grant R305D190011 from the Institute for Educational Sciences.

Contact Information:

Daniel McNeish, PO Box 871104, Arizona State University, Tempe, AZ 85287. Email: dmcneish@asu.edu

Abstract

Deciding which random effects to retain is a central decision in mixed effect models. Recent recommendations advise a maximal structure whereby all theoretically relevant random effects are retained. Nonetheless, including many random effects often leads to nonpositive definiteness. A typical remedy is to simplify the random effect structure by removing random effects or associated covariances. However, this practice is known to bias estimates of remaining covariance parameters and compromise fixed effect inferences. Cholesky decompositions frequently are suggested as an alternative and are automatically implemented in some software. Instead of Cholesky decompositions, we describe factor analytic structures as an approach to avoid nonpositive definiteness. This approach is occasionally employed in biosciences like plant breeding, but, ironically, has not been established in behavioral sciences despite the close historical connection with factor analysis in these fields. We discuss how a factor analytic structure facilitates estimation and conduct simulations to compare convergence and performance to simplifying the random effects structure or Cholesky decomposition approaches. Results show a lower rate of nonpositive definiteness with the factor analytic structure than Cholesky decomposition and suggest that factor analytic covariance structure may be useful to combating nonpositive definiteness, especially in models with many random effects.

Reducing Incidence of Nonpositive Definite Covariance Matrices in Mixed Effect Models

Multilevel and mixed effects models in psychology have become mainstream for modeling data that have a clustered structure such that people are nested within organizational units (e.g., cross-sectional data such as students in schools) or longitudinal data where repeated measures are nested within people (e.g., Grimm et al., 2016; Hox et al., 2017). A benefit of these models is that random effects are included to allow the coefficients to be specific to each higher-level unit. In cross-sectionally clustered data, this means that the relation between a predictor and the outcome is allowed to differ across organizational units to capture attributes specific to that unit. In longitudinal data, this means that growth curves are person-specific to allow for variability in change over time.

A key decision point in specifying these models is which random effects to include in the analysis. Barr et al. (2013) recommend to include the maximum number of relevant random effects and random effect correlations in models required to satisfy the theory, which has been referred to as *maximal random effect structure*. They argue that doing so maximizes the generalizability of studies across samples and can protect against adverse effects from possible model misspecifications. The advice from Barr et al. (2013) is sound theoretically, but Eager and Roy (2017) note that it may not always be pragmatic because random effect variances and correlations are difficult to estimate and models with multiple random effects are prone to nonconvergence or inadmissible solutions. Bates et al. (2018) go further stating that “the advice to ‘keep it maximal’ often creates hopelessly over-specified random effects because the number of correlation parameters to estimate rises quickly with the dimension of the random-effects vectors” (p. 18). Chief among these convergence issues is the dreaded nonpositive definite covariance matrix, which often manifests when a variance is estimated at or below 0 or when the

magnitude of the correlation between two random effects reaches or exceeds an absolute value of 1. More generally, a matrix is considered positive definite only if all its eigenvalues are positive and considered semipositive definite if all eigenvalues are non-negative (Wothke, 1993). All covariance matrices are at least semipositive definite, so nonpositive definiteness indicates that the matrix cannot be reflective of the actual relations between variables (Pinheiro & Bates, 1996). In mixed effects models, a nonpositive definite covariance matrix implies that the distribution of random effects is degenerate and that the dimensionality of the specified random effects may exceed what the data can support.[1]

Barr et al. (2013) consider confirmatory research designs common in psycholinguistics and related cognitive sciences where most or all predictors are manipulated conditions, so the number of possible random effects is often a relatively small number and dictated by the experimental design. For studies situated further towards the exploratory end of the spectrum or those with observational designs, the statistical model is not necessarily dictated by the design and multiple models with different combinations of predictors are often considered. In such cases, there is not necessarily one maximal random effect structure. Alternatively, if a researcher tried to include every possible predictor with a random effect, there would be concerns about overfitting (Raudenbush & Bryk, 2002, p. 256). The convergence difficulties described by Eager and Roy (2017) are more influential with observational data because the final model can be unduly influenced simply by what will and will not converge during model building.

[1] We are interested in the positive definiteness of the random effect covariance matrix itself here, but also note that the positive definiteness of the Hessian matrix can also be a problem (Gill & King, 2004). The Hessian matrix contains the second partial derivatives of the loglikelihood with respective to each parameter in the model and its negative inverse is used to obtain the standard errors in maximum likelihood estimation (e.g., Efron & Hinkley, 1978). If the Hessian is nonpositive definite, it cannot be inverted and the standard errors are either untrustworthy or unreliable. Reasons for nonpositive definite Hessian matrices include redundancy between variables or parameters, an overparameterized model, or a saddle point on the likelihood surface (Cudeck, Klebe, & Henly, 1993).

The problem of nonpositive definite covariance matrices is exacerbated in the common context where the sample size at the higher-level is small or moderate (Hox & McNeish, 2020). For example, in the context of modeling curvilinear growth, Diallo et al. (2014) note that the variance of quadratic slopes across people is extremely difficult to estimate with convergence often being less than 50% across simulation conditions. The number of repeated measures is closely tied to convergence rates, which fell below 30% in their study with four repeated measures. When a maximal random effect structure is attempted but a nonpositive definite matrix is encountered, a pragmatic solution is to progressively simplify the random effect structure (Barr et al., 2013; Kiernan, 2018; Miyazaki & Frank, 2006). This typically involves removing random effects or the covariances among random effects until convergence is achieved and is a popular approach in empirical studies (Baird & Maxwell, 2016; Wang et al., 2019). Though this can aid in achieving convergence and ridding oneself of a nonpositive definite covariance matrix, it can lead to covariance structure misspecifications, which are known to bias covariance parameter estimates and affect Type-I error rates for fixed effects (Diggle et al., 2002; Ferron et al., 2002; Hoffman, 2015; Weiss, 2005), which are typically inflated but can also become deflated in some circumstances (Wang et al., 2019). For instance, the simulation in Barr et al. (2013) reported that the Type-I error rates for model omitting random slopes were between 2 and 4 times the nominal rate, depending on conditions. As an example of the potential ramifications of removing random effects, Fisher, Hahn, deBruine, and Jones (2015) was retracted after the authors realized that including random slopes rendered the interaction effects of interest non-significant. Typically, removing random effects is neither substantively justified nor desirable. For instance, when fitting a quadratic growth model, a researcher might encounter a nonpositive definite covariance matrix wherein the quadratic random effect has non-zero

variance but correlates perfectly with the linear random effect. Should the researcher then remove this random effect because of the redundancy with the linear slope random effect? Doing so would imply that there are no individual differences in the curvature of the trajectories. This outcome is both at odds with the researcher's original hypothesis that individual trajectories would differ in curvature and is inconsistent with the non-zero variance estimate obtained for the quadratic effect. Moreover, the researcher would be making the dubious assumption that the perfect correlation of these random effects reflects some truth about their underlying structure in the population, rather than simply estimation difficulties with the available sample data.

In this paper, we propose that specifying a maximal random effect structure in a different way can greatly reduce the incidence of nonpositive definite covariance matrices, especially with longitudinal data containing few people or few repeated measures. In practice, a maximal random effect structure is parameterized almost exclusively via an *unrestricted structure* whereby each element in the covariance matrix is represented by a unique parameter that is explicitly estimated. Instead, we suggest using a *factor analytic* structure, which was first proposed by Jennrich and Schluchter (1986) and extended by Miyazaki and Frank (2006). The general idea of the factor analytic structure is to decompose a maximal structure into loadings and uniquenesses with a factor model (typically a one-factor model, but more factors are possible). That is, the covariance matrix is decomposed by factor analyzing the random effects within the same model. This permits retention of each element of a maximal random effect structure as in the unrestricted structure; however, in the factor analytic approach, each element is the product of other parameters rather than being a parameter that is directly estimated.

Using a factor analytic structure over an unrestricted structure has received attention in biostatistics to address estimation issues, especially in plant breeding (Piepho, 1997, 1998) and

spatial or high-dimensional mixed effect models (Smith et al., 2001; Wang, Wang, Hedeker, & Chen, 2019). The benefit of the approach is that it is more numerically stable with larger numbers of random effects while providing the same information as the traditional unrestricted structure, often with fewer parameters (Meyer, 2009). However, despite use in biostatistics and inclusion in mainstream software like SPSS and SAS Proc Mixed, applications of factor analytic covariance structures are notably absent in psychology, which is somewhat ironic given psychology's close historical connection with factor analysis. Indeed, the biostatistics literature on this topic references psychology's influence in the development of factor analysis, yet psychologists remain mostly unaware of its potential advantages in addressing issues related to nonpositive definite random effect covariance matrices.

The purpose of this paper is therefore to demonstrate the advantages of a factor analytic structure for the random effect covariance structure within mixed effects model applications in psychological research and to compare its performance to other existing methods. For expository purposes, we focus on the growth-modeling context, with the understanding that the benefits of the factor analytic structure extend to cross-sectional contexts as well. We start with a brief overview of growth modeling in the mixed effect model framework. We then provide a motivating empirical example where the estimated random effect covariance matrix is nonpositive definite, reporting the results obtained under the typical approach of assuaging nonpositive definiteness by simplifying the random effect structure. We follow with an overview of alternative ways to address convergence difficulties, with particular focus on how the factor analytic covariance structure approach can facilitate estimation and numerical stability of covariance matrices. We then reanalyze the data from our motivating example to show how the factor analytic method alleviates the nonpositive definiteness of the covariance matrix without

requiring any simplifications. We also note how several of the conclusions change when all random effects are retained in the model with this structure. To generalize this point, we conduct a simulation study to show how the factor analytic method reduces the rate of nonpositive definite matrices across multiple conditions and also improves statistical properties of estimates such as the relative bias of covariance parameters and Type-I error rates of fixed effects. We then show how the same principles apply to data with cross-sectional clustering using an empirical example where students are clustered within schools. We end with a discussion of the limitations and future directions.

Overview of Growth Models

Mixed effect models treat data univariately such that repeated measures are a single variable but are indexed by a second *Time* variable such that each repeated measure occupies a unique row and a single individual's data spans multiple rows (i.e., the data are long). *Time* is then included as an explicit predictor in the model and growth is captured by a regression coefficient associated with *Time*. These coefficients can vary across people via random effects, meaning that each person has their own person-specific growth curve.

In multilevel notation from Raudenbush and Bryk (2002), a basic unconditional linear growth model can be written as

$$\begin{aligned} y_{ti} &= \beta_{0i} + \beta_{1i} Time_{ti} + e_{ti} \\ \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \end{aligned} \tag{1}$$

The first expression is a typical regression model where the outcome y for the ith person at the tth time is equal to an intercept (β_{0i}) plus the linear slope (β_{1i}) times the tth value of *Time* plus a residual (e_{ti}) for the ith person at time t. The difference between a standard single-level regression and a growth model is that the regression coefficients (β) have i subscripts, meaning that they

vary across people. Correspondingly, each person has a unique intercept and slope which produces person-specific growth trajectories. These person-specific intercepts and slopes are directly modeled in the second and third expressions in Equation 1. The person-specific intercept is equal to the average intercept across all people (γ_{00}) plus a person-specific random effect (u_{0i}) which captures the deviation of the *i*th person's intercept from the average intercept. The person-specific slope is equal to the average slope across all people (γ_{10}) plus a person-specific random effect (u_{1i}) which captures the deviation of the *i*th person's slope from the average slope.

A common growth model applied in empirical studies is the quadratic polynomial growth model, whose popularity is derived from its ability to capture curvilinear growth while also being easy to specify due to the model being linear in the parameters (Grimm et al., 2016, p. 209). Though there are reservations about the quadratic polynomial model prioritizing ease of implementation over interpretability (Blozis, 2004; Cudeck, 1996; Cudeck & du Toit, 2002; Preacher & Hancock, 2015), we focus on the model here because it is a commonly used model whose convergence issues are well-appreciated. That is, the convergence issues we discuss in this paper appear in a variety of contexts such as cross-sectional multilevel models, cross-classified multilevel models, and longitudinal models and the quadratic growth model serves as a case in point rather than the sole interest.

The unconditional quadratic polynomial growth model can be written as

$$\begin{aligned} y_{ti} &= \beta_{0i} + \beta_{1i} Time_{ti} + \beta_{2i} Time_{ti}^2 + e_{ti} \\ \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \\ \beta_{2i} &= \gamma_{20} + u_{2i} \end{aligned} \tag{2}$$

After adding a second-order polynomial term for *Time*, the intercept is the predicted value at *Time* = 0, the linear slope is the instantaneous growth rate at *Time* = 0 (i.e., the slope of the

tangent line at the intercept), and the quadratic slope captures half the acceleration of the growth curve. Usually, the zero-point for *Time* is placed at the origin of the observation window so that the intercept represents initial status and the linear slope represents initial rate of change but other placements and interpretations are equally valid (Singer, 1998). For instance, as we will see in our demonstration, the zero-point can also be placed at the end of the observation window so that intercepts and linear slopes represent final status and final rates of change. The highest-order effect, the quadratic slope, is unaffected by this choice.

The individual values of the random effects are not explicitly estimated but instead are assumed to come from a multivariate normal distribution with a zero mean vector and an estimated covariance matrix: $\mathbf{u}_i \sim MVN\left(\mathbf{0},\mathbf{G}\right)$. Researchers can choose a structure for the random effect covariance matrix **G** where the most common structure is *unrestricted.* For the model in Equation 2 containing three random effects, an unrestricted structure would be

$\mathbf{G}=\begin{bmatrix}\tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22}\end{bmatrix}$ where the diagonal terms are variances and off-diagonal terms are covariances. All elements of the matrix are explicit parameters that are directly estimated without constraints, other than possible boundary constraints to keep variances from becoming negative. An alternative way to fit the unrestricted structure is to directly estimate the standard deviation of each random effect and the correlation between random effects and use these values to derive the variances and covariances, $\mathbf{G}=\begin{bmatrix}{\theta_{00}}^2 & & \\ \rho_{10}\theta_{11}\theta_{00} & {\theta_{11}}^2 & \\ \rho_{20}\theta_{22}\theta_{00} & \rho_{21}\theta_{22}\theta_{11} & {\theta_{22}}^2\end{bmatrix}$ where ρ designates correlations and θ represents standard deviations such that $\theta=\sqrt{\tau}$.

Importance of the Random Effect Covariance Matrix

In addition to being important for quantifying the variability in coefficients across people, proper specification of the random effect covariance matrix is an explicit assumption for proper inference in mixed effect models because it is directly related to estimation of the fixed effect standard errors. To show this, it is helpful to rewrite Equation 1 in general matrix notation from Laird and Ware (1982) such that

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \tag{3}$$

where $\mathbf{y}_i$ is a vector of scores containing person *i*'s repeated measures, $\mathbf{X}_i$ is a design matrix for the fixed effects containing person *i*’s values for the predictor variables (e.g., 1, *Time*, and $Time^2$ for quadratic growth), $\boldsymbol{\gamma}$ is a vector of fixed effect coefficients, $\mathbf{Z}_i$ is a design matrix for the random effects containing person *i*’s values of variables whose effect varies across people, $\mathbf{u}_i$ is a vector of person *i*'s random effects, and $\mathbf{e}_i$ is a vector of person *i*'s residuals such that $\mathbf{e}_i \sim MVN\left(\mathbf{0}, \mathbf{R}_i\right)$. The residual covariance matrix is usually assumed to be diagonal, i.e., $\mathbf{R}_i = diag\left(\sigma_{1i}^2, \sigma_{2i}^2, ..., \sigma_{Ti}^2\right)$, where σ_{ti}^2 might be freely estimated or estimated subject to some constraint to model heteroskedasticity as a function of time with few parameters (e.g., $\sigma_{ti}^2 = \sigma^2 \exp\left[\omega \times Time_{ti}\right]$). In other situations, the residual covariance matrix is assumed to be homoskedastic, i.e., $\mathbf{R}_i = \sigma^2\mathbf{I}_i$. For closely spaced repeated measures, serial correlation structures could also be considered (Fitzmaurice et al., 2012).

In a mixed effect model, the model-based estimate of the sampling variability of the fixed effect coefficients is calculated by

$$Cov\left(\hat{\boldsymbol{\gamma}}\right) = \left[\sum_{i=1}^{N} \mathbf{X}_i^{\mathrm{T}} \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i\right]^{-1} \tag{4}$$

where $\hat{\mathbf{V}}_i$ is the estimated marginal covariance such that $\hat{\mathbf{V}}_i = \mathbf{Z}_i \hat{\mathbf{G}} \mathbf{Z}_i^{\mathrm{T}} + \hat{\mathbf{R}}_i$. The standard errors are then calculated by taking the square root of the diagonal elements of Equation 4. To the extent that the structure for the random effect covariance matrix **G** and the residual covariance matrix $\mathbf{R}_i$ are incorrect, these misspecifications will permeate to the fixed effect standard errors and can yield incorrect p-values and lead to incorrect inferences (Laird & Ware, 1982). Wang et al. (2019) analytically show the impact that different types of misspecifications have on the fixed effect standard errors for different types of estimation methods. This represents one motivation for the maximal random effect structure: if relevant random effects are excluded or relations between random effects are not captured, researchers not only miss out on the ability to capture the variability in the coefficient across people but also increase their susceptibility to faulty inference for the fixed effects. Though corrected standard errors can be computed that are robust to covariance matrix misspecifications, these are less efficient than correctly modeling the covariance (Gurka et al., 2011), less effective with smaller samples (McNeish & Stapleton, 2016), and cannot be used in conjunction with the more sophisticated degree of freedom methods that provide optimal inference (SAS Institute Inc., 2018b, p. 6551).

Motivating Example

Consider data from a randomized trial comparing treatments for schizophrenia that originally appeared in Davis (2002). The study contains 40 male patients who are randomly assigned to one of two treatment groups such that each group has 20 patients. The Brief Psychiatric Rating Scale (BPRS) is administered to these patients to assess severity of 18 symptoms such has hostility, suspiciousness, hallucination, and grandiosity. Each symptom is scored on a 1- to 7-point scale where higher numbers indicate higher severity. The score on all 18 symptoms are then summed to yield a total score with a possible range of 18 to 126. The

BPRS is administered at baseline and then at weekly intervals for eight weeks, yielding 9 total measures. The data are time-structured such that all patients have the same number of measurement occasions. The goal of the analysis is (a) to determine if patients are still improving at the conclusion of the study and (b) to determine whether there is a difference in the patients' improvement between the two treatment groups.

Figure 1 shows the empirical trajectory plots for all 40 men by treatment group. The plots show a clear curvilinear relation with the variance appearing to decrease over time. Based on this information, we fit a model with a quadratic polynomial term for *Time* and used a heterogeneous residual structure to reflect to decreasing variability over time. Specifically, the model fit to the data was

$$\begin{aligned} BPRS_{ti} &= \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti} \\ \beta_{0i} &= \gamma_{00} + \gamma_{01}Treat_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}Treat_i + u_{1i} \\ \beta_{2i} &= \gamma_{20} + \gamma_{21}Treat_i + u_{2i} \end{aligned} \tag{5}$$

where

$$\mathbf{u}_i \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} {\theta_{00}}^2 & & \\ \rho_{10}\theta_{11}\theta_{00} & {\theta_{11}}^2 & \\ \rho_{20}\theta_{22}\theta_{00} & \rho_{21}\theta_{22}\theta_{11} & {\theta_{22}}^2 \end{bmatrix}\right) \tag{6}$$

and

$$e_{ti} \sim N\left(0, \sigma^2 \exp\left[\omega Time_{ti}\right]\right) \tag{7}$$

Equation 5 shows that BPRS scores are a function of an intercept, linear slope, and quadratic slope. *Time* was coded so that the last time-point was equal to 0 and the first time-point was -8 in order to directly explore behavior of BPRS scores at the conclusion of the study (e.g., the intercept corresponds to the average difference at the end of the study, rather than at the

beginning). The equations for each of these coefficients includes a fixed effect, a treatment effect, and a random effect to allow the coefficient to be person-specific. The random effects are assumed to be normally distributed with a zero mean vector and an unrestricted covariance matrix. In this example and throughout this paper, we focus on parameterizing the unrestricted structure using standard deviations and correlations rather than variance and covariances to facilitate identifying whether off-diagonal elements are out of bounds (Stroup et al., 2018). The residuals are assumed to be normally distributed with a heterogeneous variance that changes as a function of time.[2]

[Figure 1 about here]

We fit the model in SAS Proc Mixed with restricted maximum likelihood estimation and Kenward-Rogers degrees of freedom (Kenward & Roger, 1997) to best account for the smaller sample size (McNeish, 2017). Estimation was conducted with SAS default options which include 50 iterations each with a maximum of 150 optimization calls per iteration and convergence determined by a relative change in the Hessian matrix being less than 1e-8. Boundary constraints are also included by default to prevent variances from being below 0 and to prevent the magnitudes of correlations from exceeding 1. The results showed that the random effect covariance matrix was nonpositive definite.[3] Specifically, the linear slope variance was estimated to be 0, which led to undefined correlations between the intercept and linear slope and between the linear slope and quadratic slope. Following the common troubleshooting approach, we

---

[2] We began with a heterogeneous diagonal structure for the residuals, but because there are a fairly large number of repeated measures, this structure required 9 parameters and was not parsimonious. The structure shown in Equation 7 accomplishes the same goal with only 2 parameters and was more parsimonious based on the BIC (2462.0 vs. 2460.0).

[3] This result is unaffected by parameterizing the unrestricted structure with variances and covariances or standard deviations and correlations. We also fit the model with a homoskedastic residual structure to confirm that the issue was not attributable to the heterogenous variance structure, and the issue remained.

simplified the random effect covariance matrix by removing the quadratic random effect (the highest-order term).

The average fitted curve for each group is shown in Figure 2 and the model estimates using the simplified random effect structure are shown in Table 1. To answer the first research question about differences between groups after 8 weeks, the fixed effect of treatment is not significant ($\gamma = 5.68, t(39.8) = 1.59, p = .12$), indicating that the means of the groups are about the same at the end of the study. To answer the second research question about the rate of change at the end of treatment, the results show that the instantaneous slope of Treatment 1 is negative at 8 weeks after baseline but the *p*-value is just above the threshold for statistical significance ($\gamma = -1.15, t(173) = -1.93, p = .05$). The Treatment Group 2 instantaneous slope at Week 8 is positive and statistically significant ($\gamma = 1.35, t(173) = 2.26, p = .03$), meaning that Treatment 2 appears to have worn off and that scores are increasing at the conclusion of the study. The difference between the instantaneous slopes across Treatment Groups is significant ($\gamma = 2.50, t(173) = 2.96, p < .01$). The quadratic slope was positive and statistically significant both for Treatment 1 ($\gamma = 0.20, t(278) = 2.28, p < .01$) and Treatment 2 ($\gamma = 0.41, t(278) = 5.88, p < .01$) and there was a significant difference between the quadratic slopes such that the growth trajectory for Treatment 1 showed more significantly more curvature ($\gamma = 0.21, t(278) = 2.16, p = .03$). Overall, we conclude that Treatment 1 is no longer significantly improving at the conclusion of the study whereas Treatment 2 has reversed course and BPRS scores are increasing at the conclusion of the study, which is attributable to the significantly greater bend in the growth trajectory.

[Table 1 about here]

[Figure 2 about here]

Approaches to Improve Convergence

The previous section noted showed an example of how convergence difficulties can prompt a researcher to respecify their model from what was originally intended. While simplification of the random effects structure is typical, several alternative approaches to model re-specification are also common to varying degrees. We review these here prior to exploring our preferred solution – the factor analytic structure – in greater detail.

Simple Rescaling

One potential cause of estimation difficulties can be differences in the scales of the predictors. For instance, when higher-order polynomials are added to growth models, the values of the second-order *Time*^{2} variable have a wider range than the first-order *Time* variable. For instance, in the BPRS data, the fixed effects are in units of *Time* and *Time*^{2}, which range from -8 to 0 and 0 to 64, respectively. In turn, the variance and covariance parameters associated with the random effects of *Time* and *Time*^{2} are likely to be quite different in magnitude, and this makes estimation difficult (Cheng et al., 2010). Thus, rescaling predictors to have more similar ranges is one way to potentially improve convergence to a proper solution. For instance, in the BPRS data, we could have rescaled *Time* to range from -1 to 0 such that a one-unit change corresponds to the entire window of observations. In turn, *Time*^{2} would range from 0 to 1. This change of scale might help to stabilize estimation, improving the likelihood of convergence; however, it is less likely to help to avoid nonpositive definite solutions.

Centering

Centering *Time* by subtracting its mean is one specific way to rescale the predictors to prevent higher-order polynomial terms from growing much more quickly than lower-order terms

(i.e., the range of $Time^2$ would only be 0 to 20.25 instead of 0 to 64 if *Time* were mean-centered; Schuster & von Eye, 1998). Unlike many other potential rescaling options, centering also has the advantage that it greatly reduces collinearity between polynomial terms or removes it entirely in time-structured data (Hedeker & Gibbons, 2006 p. 86), potentially further improving convergence and reducing the likelihood of a nonpositive definite solution.

Centering is not always optimal for interpretation as it necessarily shifts the intercept to the midpoint of the observation window, seldom a location of substantive interest (Raudenbush, 2001). For instance, in the BPRS data, *Time* was meaningfully coded to be zero at the end of the observation window to facilitate evaluating individual and group differences at the end of the trial, so centering would affect the ability to address the original research question as a direct parameter in the model (the endpoint could be tested with any centering scheme through linear combinations of parameters, however).

Orthogonal Polynomial Model

Orthogonal polynomials originating from analysis of variance have been suggested as another method that rescales *Time* (Hedeker & Gibbons, 2006 p. 84). Applying this method requires some data preprocessing to transform the time values into orthogonal polynomial values (e.g., Hedeker & Gibbons, 2006 p. 88), but it has the benefit of putting the polynomial terms for *Time* variables onto the same scale and removing the correlation between them. Specifically, if $\mathbf{T}$ is the matrix of time values (e.g., intercept, *Time*, and $Time^2$ for a quadratic growth model), the orthogonal polynomial value matrix is calculated by $\mathbf{T}\left(\mathbf{S}^{\mathrm{T}}\right)^{-1}$ where $\mathbf{S}$ is the Cholesky factor of the symmetric matrix, $\mathbf{T}^{\mathrm{T}}\mathbf{T}$. A benefit of orthogonal polynomials is that nonpositive definiteness related to random effect correlations greater than 1 can be alleviated, although lower boundary issues pertaining to random effect variances near 0 can persist. Nonetheless, similar to centering,

interpretation can be difficult. Consequently, Biesanz, Deeb-Sossa, Papadakis, Bollen, and Curran (2004) suggest that orthogonal polynomial models are best understood graphically but ultimately caution against their use (although the coefficients can be translated to correspond to the original time metric, Hedeker & Gibbons, 2006, p. 91).

Cholesky Covariance Structure

Whereas centering and orthogonal polynomials are potentially useful when trying to address non-positive definiteness in models that include random effects of higher order terms (e.g, $Time^2$), these solutions are not as applicable for models that do not include higher-order polynomial terms. An alternative and fully general strategy that is sometimes suggested is to reparameterize the random effects covariance matrix via a Cholesky decomposition.

As noted in the introduction, not every symmetric matric conforms to the properties of a covariance matrix. As a result, direct estimation of an unrestricted structure may result in a nonpositive definite and therefore inadmissible estimate of the random effects covariance matrix, unless constraints are applied to ensure positive semi-definiteness (Pourahmadi, 1999). Reparameterizing the covariance matrix in terms of a Cholesky decomposition, however, ensures at least semipositive definiteness without requiring constraints on the parameters.

The Cholesky decomposition (**L**) is the matrix square root of **G** with $r\left(r+1\right)/2$ parameters where $\mathbf{G}=\mathbf{L}\mathbf{L}^{\mathrm{T}}$.[4] Rather than estimating **G** directly, the goal is to estimate the elements of **L** and then multiply **L** by itself (transposed) to create **G**. The benefit is that none of the elements in **L** need to be constrained but $\mathbf{L}\mathbf{L}^{\mathrm{T}}$ is guaranteed to be at least semipositive definite, similar to how the square of any real number (positive or negative) cannot be negative.

[4] There is also a modified Cholesky factorization that estimates the matrix square root of the *correlation* matrix and an *r*-dimensional diagonal matrix of variances ($\boldsymbol{\Sigma}$) such that $\mathbf{G}=\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^{\mathrm{T}}$ (Daniels & Zhao, 2003).

Structuring of the covariance matrix with a Cholesky decomposition can still produce a nonpositive definite matrix where one or more diagonal terms are exactly zero, however, so a Cholesky decomposition only guarantees *semi*-positive definiteness rather than positive definiteness. A Cholesky decomposition for the random effect covariance matrix is commonplace in software (e.g., Hedeker & Gibbons, 1996; Bates, Mächler, Bolker, & Walker, 2014) because it increases stability of estimation, is less susceptible to rounding problems, and is computationally expedient (Lindstrom & Bates, 1988, SAS Institute Inc., 2018a, p. 3730)._The idea of making variance scale magnitudes more congruent to ease numerical stability of estimation has been noted previously. Kohli et al. (2019) discuss its effectiveness (and often, its necessity) in piecewise models with random knots. Additionally, McNeish, Dumas, and Grimm (2019) describe reparameterizing nonlinear models to reduce scale magnitude differences between linear parameters and those that appear in exponents.

Although the Cholesky decomposition is one general solution to the problem of nonpositive definiteness, it may still encounter estimation difficulties and fail to produce a converged solution. A different covariance structure approach – namely, the factor analytic covariance structure – shares some of the desirable properties of a Cholesky decomposition, but also has some unique properties of its own. However, the method of using a factor analytic covariance structure has not been widely recognized among psychologists and, as we argue, deserves greater consideration. The next section covers details of the factor analytic covariance structure and how it is suited to assistance with convergence issues like those observed in the BPRS data.

The Factor Analytic Covariance Structure

Factor analysis is well-known in psychology as a method to test hypotheses involving latent variables (Brown, 2015; Mulaik, 2009). Though this is a common substantive interest, factor analysis at its core is a mathematical decomposition of the covariances between variables (Meyer, 2009). That is, when psychologists test global fit of factor analysis models with a likelihood ratio test, the null hypothesis posits that the covariance matrix implied by the model is equal to the observed covariance matrix of the variables. Factor analysis ultimately boils down to reproducing a covariance matrix based on hypothesized common factors that explain relations between observed variables. This property of factor analysis makes it potentially useful as computational tool for decomposing a random effect covariance matrix, even if there is not necessarily theoretical interest in common factors underlying the random effects (Jennrich & Schluchter, 1986; Wolfinger, 1996).

To clarify, the typical unrestricted structure as featured in Equation 6 directly estimates the standard deviation and correlations (or the variances and covariances) between random effects. As a conceptual path diagram, this can be represented by the top panel of Figure 3. Each variance or covariance is an explicit parameter, and these parameters are estimated directly. This results in $r\left(r+1\right)/2$ unique parameters being estimated for *r* random effects. On the other hand, the factor analytic approach decomposes the covariance between the random effects with, in this case, a one-factor model, whose path diagram is shown in the bottom panel of Figure 3. The factor analytic structure features no direct estimates of random effect variances or covariances. Instead, it estimates loadings (λ) and uniquenesses (*d*) to decompose the covariances between the random effects. These loadings and uniquenesses are then combined to yield each element of the random effect covariance matrix such that $\mathbf{G}=\mathbf{\Lambda}\mathbf{\Lambda}^{\mathrm{T}}+\mathbf{D}$ where $\mathbf{\Lambda}$ is a $r\times 1$ vector of factor loadings and $\mathbf{D}$ is a $r\times r$ diagonal matrix of uniquenesses (readers may recognize this as the

model-implied covariance from factor analysis when the factor covariance is an identity matrix). Note that the uniquenesses in **D** are not the variance of the residuals from the growth model discussed earlier in $\mathbf{R}_i$ (which reside at Level 1) but instead are parameters associated with the variability in the random effects that are not accounted for by the loadings in $\mathbf{\Lambda}$ (which reside at Level 2). The factor analytic structure requires 2*r* parameters (*r* loadings and *r* uniquenesses) to fill the complete random effect covariance matrix.[5]

[Figure 3 about here]

To make this explicit, consider the quadratic polynomial growth model from the BPRS data where r = 3. The factor analytic approach would estimate 3 loadings and 3 uniquenesses which would be combined in the following way to populate the random effect covariance matrix

$$\mathbf{G} = \mathbf{\Lambda\Lambda}^{\mathrm{T}} + \mathbf{D} = \begin{bmatrix} \lambda_0^2 + d_0 & & \\ \lambda_0\lambda_1 & \lambda_1^2 + d_1 & \\ \lambda_0\lambda_2 & \lambda_1\lambda_2 & \lambda_2^2 + d_2 \end{bmatrix} \tag{8}$$

Each variance is equal to the loading for that random effect squared plus the uniqueness for that random effect. Each covariance is equal to the product of the loadings for the two random effects that are involved. In this way, each element still takes on a unique numerical value as in the unrestricted structure; however, these elements are not directly estimated but rather are formed from estimates of a factor analytic decomposition of the random effects. Just like how we can parameterize the covariance as a function of standard deviations and correlations, the factor analytic structure parametrizes the matrix as a function of loadings and uniqueness. The end result for each element is still a variance or a covariance, but the parameters estimated and

[5] The dimensions listed here represent a factor analytic structure with one factor. As we discuss later, it is possible to use multiple factors, in which case the dimension of $\mathbf{\Lambda}$ would be $r \times q$ and the number of parameters would be $rq + r$ for q the number of factors.

combined to arrive at that variance or covariance are structured differently. The process of decomposing the random effect covariance matrix with a factor model would seem to make the estimation process more complex since it is a roundabout way of arriving at the same information that is directly estimated by an unrestricted structure but it has three benefits that can help to avoid nonpositive definiteness.

First, the number of parameters required for a one-factor structure (2*r*) scales more slowly than the number of parameters needed for an unrestricted structure or a Cholesky structure ($r\left(r+1\right)/2$). With 3 random effects, the minimum for which the factor analytic structure would typically be identified, the number of parameters is 6 and is the same for a factor analytic structure and an unrestricted model.[6] With more than 3 random effects, however, the one-factor structure has fewer parameters, reducing the dimensionality of the matrix while continuing to model variances and covariances of all random effects (Vermunt, 2007, p. 142). For instance, a model with 5 random effects would require $\left(5\times 6\right)/2=15$ covariance parameters with an unrestricted structure but only $2\times 5=10$ covariance parameters with a factor analytic structure. If the data cannot support as many covarying dimensions as desired with an unrestricted structure, a factor analytic structure can reduce the number of dimensions without requiring that any random effects necessarily be removed from the model.

Second, “0” estimates with the factor analytic structure do not necessarily result in nonpositive definiteness as they can with an unrestricted structure (Meyer, 2008). For instance, if the estimate of d_1 in Equation 8 were 0, this would not automatically imply a nonpositive definite covariance matrix. With an unrestricted structure where variances or standard deviations

[6] With 2 or fewer random effects the one-factor factor analytic structure would be under-identified, having more parameters than the unrestricted structure, and this over-parameterization would typically be signaled by a non-positive definite Hessian matrix.

are directly estimated, a 0 on the main diagonal results in nonpositive definiteness. With a factor analytic structure, a 0 estimate for d_1 would indicate that squaring λ_1 can alone reproduce the variance for the random effect. **D** would no longer be full rank, but that is not prohibitive for convergence because the ultimate goal is that **G** is positive definite. In other words, $\mathbf{\Lambda}$ and **D** are simply a method to populate **G** so there is more flexibility in what constitutes an in-bounds estimate compared to an unrestricted structure. This property can be useful with smaller samples or many random effects where uncertainty is higher because imprecision that can result in nonpositive definiteness with an unrestricted structure does not necessarily have the same adverse effect for a factor analytic structure (Meyer & Kirkpatrick, 2005).

Third, and similar to a Cholesky structure, the factor model operates on the square root matrix of the covariance matrix (Oman, 1991), which has benefits related to scale magnitude differences between parameters (Hedeker & Mermelstein, 1998). That is, parameters whose variables are on scales with different magnitude makes estimation of unrestricted covariance matrices difficult (Kiernan, 2018) because of the way that parameters are typically updated across iterations using the gradient vector and the Hessian matrix of the likelihood. Similar to centering or orthogonal polynomial approaches, working on the square root matrix makes parameters more stable to estimate (Cheng et al., 2010; Gibbons & Bock, 1987).

In the next section, we reanalyze the motivating data with a factor analytic approach to show how it can reduce convergence issues. Afterward, we carry out a simulation study to quantify its potential benefits to expel nonpositive definite random effect covariance matrices.

Reanalysis of Motivating Example

We return to the motivating BPRS data we discussed earlier. We fit the same model as in Equation 5 but this time using a factor analytic structure for the random effect covariance matrix such that

$$\mathbf{u}_i \sim MVN\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}\lambda_0^2+d_0 & & \\ \lambda_0\lambda_1 & \lambda_1^2+d_1 & \\ \lambda_0\lambda_2 & \lambda_1\lambda_2 & \lambda_2^2+d_2\end{bmatrix}\right) \tag{9}$$

The residual structure is again heteroskedastic as a function of time as in Equation 7. We fit the model in SAS Proc Mixed using restricted maximum likelihood estimation with Kenward-Rogers degrees of freedom with the SAS default convergence options; in SAS, a one-factor structure can be requested in the RANDOM statement by specifying TYPE = FA(1).

Whereas the model previously had issues with nonpositive definiteness using the unrestricted structure in Equation 6, with a factor analytic covariance structure, the model converged without issue. The average fitted trajectories are identical to those shown in Figure 2 and estimates for the model with a factor analytic covariance structure are shown in Table 2. The factor analytic structure does not directly estimate the individual elements of the covariance structure, so some calculation is needed to transform the estimates in Table 2 into the random effect covariance matrix:

$$\mathbf{G} = \begin{bmatrix}\lambda_0^2+d_0 & & \\ \lambda_0\lambda_1 & \lambda_1^2+d_1 & \\ \lambda_0\lambda_2 & \lambda_1\lambda_2 & \lambda_2^2+d_2\end{bmatrix} = \begin{bmatrix}6.95^2+83.14 & & \\ 6.95\times 2.16 & 2.16^2+0 & \\ 6.95\times 0.21 & 2.16\times 0.21 & 0.21^2+0.03\end{bmatrix} = \begin{bmatrix}131.44 & & \\ 15.01 & 4.67 & \\ 1.46 & 0.45 & 0.07\end{bmatrix}$$

For ease of reference, the corresponding correlation matrix is $\begin{bmatrix}1 & & \\ .61 & 1 & \\ .48 & .79 & 1\end{bmatrix}$. These estimates reveal some meaningful information that is omitted when using an unrestricted structure, such being able to quantify the quadratic slope variance and the correlations between the random

effects. Not being forced to constrain particular elements in order to achieve convergence also leads to a significant improvement in the restricted loglikelihood relative to the simplified model in Table 1, $\chi^2(3)=13.0, p<.01$ and an improvement in the BIC (2452.7 for the factor analytic structure vs. 2460.0 for the two random effect unrestricted structure).

[Table 2 about here]

Being able to model all possible elements in the random effect covariance matrix impacts the fixed effect standard errors and consequently, the *p*-values associated with the fixed effects. Despite the fact that the fixed effect point estimates are extremely close between covariance structures in Table 2, the standard errors and *p*-values differ. As a result, many of the conclusions related to the second research question presented earlier no longer hold. For instance, the instantaneous slope at Week 8 in Treatment Group 1 is now clearly not statistically significant ($\gamma=-1.17, t(36)=-1.75, p=.09$), making it clearer that growth has leveled off by Week 8 and that this group is no longer improving at the study's conclusion. The *p*-value for the instantaneous slope of Treatment Group 2 at Week 8 now slightly exceeds .05 ($\gamma=1.35, t(36)=2.02, p=.051$), dampening possible evidence that this group is getting worse at the conclusion of the study. The difference in the quadratic slopes is also no longer statistically significant ($\gamma=-0.22, t(40.1)=1.72, p=.09$), indicating that there is no difference in the curvature of the growth trajectories between treatments.

Essentially, all of the previously significant effects related to the second research questions are no longer significant when all elements of the random effect covariance matrix are retained with the factor analytic structure. Of course, this is a single empirical example and it is impossible to determine which set of results is actually more accurate. To test this more

rigorously, we perform a simulation in the next section to compare the performance of the unrestricted and factor analytic structures for the random effect covariance matrix.

Simulation

Simulation Design

To determine if the findings from the BPRS example hold more generally and to compare performance of different approaches to addressing nonpositive definiteness, we performed a Monte Carlo simulation study. There is an extensive literature surrounding Cholesky decompositions of covariance matrices to make computation more expedient or to allow for unconstrained estimation (Pinheiro & Bates, 1996; Elzo, 1996) in addition to related decompositions for specifying prior distributions in Bayesian analyses (Barnard, McCulloch, & Meng, 2000; Liu, Zhang, & Grimm, 2016). These issues were salient in years past when computation time for random effects models was more problematic, but there has been very little exploration related to the *practical* question of whether employing different approaches ultimately leads to improved convergence and performance compared to directly estimating the elements of an unrestricted covariance matrix (for exceptions that include practical comparisons as secondary interests, see van der Elst et al., 2016 and Lu & Mehrotra, 2010).

We generated data from a two-group quadratic growth model whose population values were taken from rounding the BPRS estimates presented previously. We simplified the residual covariance matrix to be a homogeneous diagonal rather than heteroskedastic. The statistical model used to generate data is shown in Equations 10 and 11.

$$\begin{aligned} y_{ti} &= \beta_{0i} + \beta_{1i} Time_{ti} + \beta_{2i} Time_{ti}^2 + e_{ti} \\ \beta_{0i} &= 30 + 5.0\left(Treat_i\right) + u_{0i} \\ \beta_{1i} &= -1.0 + 2.35\left(Treat_i\right) + u_{1i} \\ \beta_{2i} &= 0.20 + 0.20\left(Treat_i\right) + u_{2i} \end{aligned} \tag{10}$$

$$\mathbf{u}_i \sim MVN\left(\begin{bmatrix}0\\0\\0\end{bmatrix},\begin{bmatrix}130 & & \\ 15 & 5 & \\ 1.5 & 0.50 & 0.07\end{bmatrix}\right), \quad \mathbf{e}_i \sim N\left(0,15\right) \tag{11}$$

Measurement occasions were generated to be time-structured such that each person had the same number and timing of measurements. Diallo et al. (2014) noted that the number of repeated measures was closely related to convergence in quadratic growth models, so we included three repeated measure conditions: 4, 6, and 8. We included two variations of the 4 repeated measure conditions with different timing. In the “narrow” 4 repeated measure condition, timing corresponded to the final 4 time-points of the 8 repeated measure condition (i.e., time was coded -3, -2, -1, and 0). In the “wide” 4 repeated measure condition, the window of observation was equal to the 8 repeated measure condition but with half as many measurement occasion (i.e., time was coded -6, -4, -2, 0). We also included three different sample sizes: 50, 100, and 200. We focus on the smaller end of the sample size spectrum because this is where issues associated with nonpositive definiteness are most common. For each combination of sample size and repeated measures, we created 1000 datasets from the data generation model using SAS Proc IML. To each generated dataset, we fit the model in Equation 10 in SAS Proc Mixed using restricted maximum likelihood estimation and the default convergence and boundary constraint options. SAS files used to conduct the simulation are provided on the first author’s Open Science Framework page (https://osf.io/xhfqw).

To each generated dataset, we fit models with 4 different parameterizations:

1. An unrestricted random effect covariance structure with a standard deviation-correlation parameterization

2. Re-expressing *Time* and *Time*^{2} via orthogonal polynomials and using an unrestricted random effect covariance structure with a standard deviation-correlation parameterization
3. An unrestricted random effect covariance structure parameterized by a Cholesky decomposition
4. A factor analytic random effect covariance structure based on one-factor

As in the analysis of the example data, for replications that did not converge using the unrestricted structure with a standard deviation-correlation parameterization, we fit a model that removed the random effect for the quadratic slope so that the random effect covariance matrix was 2×2 and only included the intercept standard deviation, the linear slope standard deviation, and the correlation between the two to explore the statistical properties of removing random effects when it may not be warranted. Refitting the model with fewer random effects was only done for the first parameterization in the above list.

Outcomes

The first outcome of interest was the convergence of the fitted models. Nonpositive definiteness of the estimated random effect covariance matrix was the main interest and the main source of convergence issues, but we also counted other convergence issues such as a nonpositive definite Hessian matrix or an infinite likelihood.

Second, we looked at the relative bias of the random effect standard deviations and random effect correlations in replications that were able to converge for each type of covariance structure. This outcome was of interest because the ability to achieve convergence is necessary but not sufficient for recommending use of a particular method. A method that converges but yields poor estimates is not necessarily an improvement, so this outcome will help address

whether the alternative covariance structures increase, decrease, or has no effect on the quality of the covariance structure estimates. The random effect covariance population values vary widely in their magnitude, which can affect the comparability of relative bias across parameters. For this reason, we report the relative bias of the random effect correlations instead in attempt to somewhat standardize the relative bias values across the different off-diagonal elements. A similar rationale is behind the choice to present relative bias of standard deviations.

Third, we will report the relative bias of the fixed effect estimates and the coverage of the 95% confidence interval for each of the fixed effects. Relative bias of the fixed effect estimates is not expected to be affected by the specification of the random effect covariance matrix with continuous outcomes (Liang & Zeger, 1986) but we will inspect this metric as a precaution. However, the standard errors of the fixed effects may be adversely affected. To explore this possibility, we inspect the proportion of replications in which the population value falls within the estimated 95% confidence interval for each fixed effect parameter. If there is minimal relative bias of the fixed effects, this measure becomes an assessment of the standard errors. If the standard errors are appropriately estimated, the confidence interval coverage rate should be near 95% with values between 93% and 97% typically being considered reasonably close to 95% (Bradley, 1978). If the standard errors are too small, the coverage rate will be below 93%; if the standard errors are too large, the coverage rate will exceed 97%. To maintain consistency with the approach commonly taken in empirical studies, we will report these metrics for

1. The converged replications of the unrestricted random effect covariance structure using a standard deviation-correlation parameterization
2. The model with only two covarying random effects for the replications where the unrestricted matrix did not converge

3. The converged replications of the factor analytic structure
4. The converged replications of the Cholesky decomposition

Results

Convergence Rates. Table 3 shows the percentage of replications that converged across each of the simulation conditions. As noted by Diallo et al. (2014), convergence was highly influenced by the number of repeated measures in the data with a higher number of repeated measures being associated with higher convergence rates. Sample size was also related to convergence rates such that convergence generally improved with larger samples, though this effect was weaker than the effect of the number of repeated measures.

[Table 3 about here]

Beyond these general trends that have been noted in previous studies, the salient finding in Table 3 is the difference in convergence between using an unrestricted structure and alternative parameterizations of *Time* or the covariance structure. With 8 repeated measures, the factor analytic structure converged in nearly every replication, with orthogonal polynomials and the Cholesky structure not far behind (mid 90s to 100 percent convergence). On the other hand, the unrestricted structure converged only about two-thirds of the time with 8 repeated measures. With 6 repeated measures, the convergence rate of the factor analytic structure was not perfect but remained noticeably higher (high 70s to mid 90s, depending on sample size) compared to either the orthogonal polynomial or Cholesky structure (high 50s to low 80s) or the unrestricted structure (mid 20s to high 30s). With 4 narrowly spaced repeated measures, the convergence of the factor analytic structure was not great (mid 40s to high 50s) but nonetheless was a moderate improvement over orthogonal polynomials or a Cholesky structure (low to mid 40s) and all three

were a vast improvement over a unrestricted structure in the same conditions (whose convergence did not exceed 20%).

As anticipated by the previously reviewed biostatistics literature, a factor analytic structure clearly demonstrates fewer obstacles to convergence compared to either unrestricted structure or an orthogonal polynomial model across conditions, especially with moderate sample size and a moderate number of repeated measures. Though the ability to increase convergence is useful, the utility of such an increase is only relevant if the quality of the estimates is also satisfactory. The next subsection reports on the bias of the covariance parameters across simulation conditions to explore whether the improvement in convergence is concomitant with quality parameter estimates.

Covariance Parameter Estimate Relative Bias. In this section, we look at the ability of the unrestricted, Cholesky structure, and factor analytic structures to recover the population values for the random effect covariance matrix. We do not report results for orthogonal polynomials, as the estimates are on a different scale than other methods due to the different coding of time. To keep results succinct, we only present the narrow 4 repeated measure condition where the greatest differences were observed.

Table 4 shows the relative bias of the random effect standard deviations and random effect correlations across simulation conditions. The relative bias values in Table 4 for each covariance structure are only calculated for replications that converged, so the number of replications is not constant across columns. For instance, the UNR column for N = 50 and RM = 4 is based on 140 replications (14% of the 1000 possible replications that converged as reported in Table 3) whereas the FA column for N =50 and RM = 4 is based on 440 replications (44% of the 1000 possible replications that converged as reported in Table 3).

[Table 4 about here]

A salient finding in Table 4 is that the relative bias for the random effect standard deviations with a factor analytic structure were generally equal to or lower than those from an unrestricted structure or Cholesky structures. This is notable because the factor analytic structure was able to converge more often than an unrestricted or a Cholesky structure *and* provided estimates that were closer to the population standard deviation values, on average. No detriment was observed in the random effect standard deviations even when factoring in the additional replications that were unable to converge with the unrestricted structure.

The relative bias of the random effect correlations does not strongly support one method being superior to the other because the best method is split across conditions. The relative bias of the correlation between random intercepts and random quadratic slopes was typically worse for the factor analytic structure. However, this finding appeared to be related to convergence. The scale magnitudes of the intercept and the quadratic slope were the most discrepant, meaning that this parameter was the most difficult to estimate. With the unrestricted structure, difficulty with this parameter may simply have triggered a nonconvergent replication. Conversely, the factor analytic structure would have been more likely to provide an estimate for this parameter, but the difficulty may have manifested in more biased estimates rather than nonconvergence. When considering only replications that converged, the factor analytic structure was always better than the unrestricted structure for estimating random effect correlations with 4 repeated measures, about the same as the unrestricted structure with 6 repeated measures, and closer to (but still slightly worse than) the unrestricted structure with 8 repeated measures.

So far, the factor analytic structure has shown that it improves convergence while moderately increasing the quality of the covariance parameter estimates compared to the

unrestricted structure. The next subsection inspects the confidence interval coverage rates when employing each of the different covariance structures.

Fixed Effect Relative Bias and Confidence Interval Coverage Rates. As anticipated, the relative bias of the fixed effect estimates was negligible across all conditions and was not different across the different covariance structure conditions. We do not report this information in the interest of brevity but do note that this signifies that any issues with confidence interval coverage rates is solely attributable to inaccuracies in the standard errors.

Table 5 shows the 95% confidence interval coverage rates for each fixed effect across the simulation conditions. For succinctness and comparability, we reported on the same conditions as Table 4. Reported values are the coverage of the 95% confidence interval, so coverage is expected to be near 5% if standard errors are accurately estimated. Though none of the fixed effects in the present data generation model are null, these coverage rates provide a good indication of what we would expect the Type-I error rate to be for a null effect under these conditions. If, for a null effect, (1 – Coverage Rate) > 0.05, then the Type I error rate is elevated above the nominal level. Thus, coverage rates under 95% for non-null effects suggest a corresponding elevation in Type-I errors for null effects without having to explicitly include null effect in the data generation model. The unrestricted structure (the UNR column), Cholesky structure (the CD column), and the factor analytic structure (the FA column) display the coverage rates for the replications that converged. The version of the model that omits the quadratic slope random effect (the UNR2 column) includes replications for which the full unrestricted structure matrix did *not* converge. This was done to mimic the common protocol of simplifying the random effect covariance matrix when the full unrestricted model fails to converge.

[Table 5 about here]

From Table 5, the unrestricted structure yields reasonable coverage rates across conditions, as would be expected since this model directly corresponds to the data generation model. However, as noted in Table 3, this model can have convergence issues and may not reliably converge to a solution in some conditions. In such cases where the random effect covariance needs to be simplified to achieve convergence, the confidence intervals were too narrow for some parameters (coverage rates below 93%), particularly when there are more repeated measures or the sample size is larger. This suggests that the Type-I error rate for null effects would also be elevated. On the other hand, the factor analytic and Cholesky structures consistently produced fixed effect coverage rates between 94 and 97%, suggesting that Type-I error rates would also be well controlled.

Summary. Alternative parameterizations such as orthogonal polynomials, Cholesky decompositions, or a factor analytic structure are not a cure-all solution to all estimation issues related to nonpositive definiteness. However, they vastly boost convergence, improve estimation accuracy for the random effect variances, do not adversely impact estimation of the random effect correlations, and generate accurate confidence intervals for fixed effects across sample size and repeated measure conditions. The factor analytic and Cholesky structures do not alter the ability to get unique estimates for all elements of the random effect covariance matrix – they merely change how those estimates are obtained and, in doing so, improve many aspects of estimation quality. Therefore, if nonconvergence issues related to nonpositive definiteness are encountered, these reparameterizations should be considered prior to removing random effects or constraining random effect correlations to 0.

Extension to Modeling Cross-Sectionally Clustered Data

Though our focus has been on growth models because issues with estimating the variance of quadratic slopes is widely noted, the convergence issues we discuss similarly extend to cross-sectional data where people are nested within higher-level units. Consider the seminal example provided in Raudenbush and Bryk (2002). The example features a subset of the 1982 High School and Beyond data set where the outcome variable is Mathematics Achievement for 6,774 students within 160 different schools. We use a slightly different subsample than used in Raudenbush and Bryk (2002) to highlight issues central to this paper[7]; for the 97 schools with fewer than 50 students, we used all students as used in Raudenbush and Bryk (2002). For the 63 schools with more than 50 students, we randomly sampled 50 students. This did not affect the unbalanced structure of the data – each school had a potentially different number of students and the median number of students per school (which was 47) was unaffected. The data include three within-school predictors (Sex, Minority Status, Socioeconomic Status [SES]) and two between-school predictors (Private School Status and School Size Divided by 100).

We will follow a top-down model building strategy and fit a model featuring all 3 student-level predictors with School Size forming a cross-level interaction with Female and Private School Status forming a cross-level interaction with both Minority Status and SES. All within-school predictors were group-mean centered and the school mean of each within-school predictor was also included to decompose effects into orthogonal within- and between-school components. The school means of the binary predictors were also grand-mean centered to

[7] The exact subsample from Raudenbush and Bryk (2002) converges using group-mean center for the model of interest, but not using grand-mean centering. Grand-mean centering can produce conflated estimates when random slopes are present (Rights et al., 2020), so we use a slightly different subsample to demonstrate the issue with the unconflated group-mean centering approach.

preserve the interpretation of the intercept. Further, we included random intercepts and random slopes for all 3 within-school predictors. In other words, the model would be

$$
\begin{aligned}
&MathAch_{ij} = \beta_{0j} + \beta_{1j}\left(Female_{ij} - \overline{Female}_j\right) + \beta_{2j}\left(Minority_{ij} - \overline{Minority}_j\right) + \beta_3\left(SES_{ij} - \overline{SES}_j\right) + e_{ij} \\
&\beta_{0j} = \gamma_{00} + \gamma_{01}Private_j + \gamma_{02}\left(Size_j - \overline{Size}\right) + \\
&\qquad \gamma_{03}\left(\overline{Female}_j - \overline{Female}\right) + \gamma_{04}\left(\overline{Minority}_j - \overline{Minority}\right) + \gamma_{05}\overline{SES}_j + u_{0j} \\
&\beta_{1j} = \gamma_{10} + \gamma_{12}\left(Size_j - \overline{Size}\right) + u_{1j} \\
&\beta_{2j} = \gamma_{20} + \gamma_{21}Private_j + u_{2j} \\
&\beta_{3j} = \gamma_{30} + \gamma_{31}Private_j + u_{3j}
\end{aligned}
$$

(12)

Where *i* indexes students and *j* indexes schools. We fit the model in SAS Proc Mixed with restricted maximum likelihood and between-within degrees of freedom because the sample size is large. We also calculated contextual effects for Female, Minority Status, and SES by subtracting the within-school coefficient from the between-school coefficient (e.g., $\gamma_{03} - \gamma_{10}$ for Female).

We initially used a 4×4 unrestricted structure for the random effect covariance matrix parameterized with standard deviations and correlations,

$$
\mathbf{u}_j \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} {\theta_{00}}^2 & & & \\ \rho_{10}\theta_{11}\theta_{00} & {\theta_{11}}^2 & & \\ \rho_{20}\theta_{22}\theta_{00} & \rho_{21}\theta_{22}\theta_{111} & {\theta_{22}}^2 & \\ \rho_{30}\theta_{33}\theta_{00} & \rho_{31}\theta_{33}\theta_{11} & \rho_{32}\theta_{33}\theta_{22} & {\theta_{33}}^2 \end{bmatrix}\right) \tag{13}
$$

However, this model was unable to converge due to a nonpositive definite covariance matrix attributable to the random effect correlation between Female and SES being inadmissible. Refitting the model after removing the Female random effect also resulted in a nonpositive definite random effect covariance matrix. Either removing the SES random effect (BIC = 43,623) or both the Female and SES random effects (BIC = 43,610) were able to converge.

When removing both random effects as suggested by the BIC, the contextual effect for Minority Status ($\gamma_{04} - \gamma_{20} = 1.16, t(6608) = 1.91, p = .057$) is not significant. However, if only the SES random effect is removed the Minority contextual effect ($\gamma_{04} - \gamma_{20} = 1.28, t(6608) = 2.09, p = .036$) is significant. This highlights the increased arbitrariness of the simplification approach with cross-sectionally clustered data – it is not obvious which random slopes may be more or less imperative to retain in the model and inference for borderline effects can change depending on the choice. Further, these choices are not made based on theory but rather upon what will or will not converge. An additional caveat is that information criteria like BIC are a common guide but have been shown to be untrustworthy for accurately selecting among models with different covariance structures and BIC often prefers overly simple structures (Ferron et al., 2002; Gomez, Schaalje, & Fellingham, 2005; Keselman et al., 1999). Estimates and standard errors from these two models are provided in the left and center columns of Table 6.

[Table 6 about here]

The orthogonal polynomial approach was inapplicable to this example, owing to the lack of higher-order polynomial terms, leaving a Cholesky or factor analytic structure as possible methods to include random effects on all four within-school predictors. A Cholesky structure would specify that

$$\mathbf{u}_j \sim MVN\left(\mathbf{0}, \mathbf{L}\mathbf{L}^{\mathrm{T}}\right) \tag{14}$$

where

$$\mathbf{L} = \begin{bmatrix} l_{00} & 0 & 0 & 0 \\ l_{10} & l_{11} & 0 & 0 \\ l_{20} & l_{21} & l_{22} & 0 \\ l_{30} & l_{31} & l_{32} & l_{33} \end{bmatrix} \tag{15}$$

and a factor analytic structure for the random effect covariance matrix would specify that

$$\mathbf{u}_j \sim MVN\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}\lambda_0^2+d_0 & & & \\ \lambda_1\lambda_0 & \lambda_1^2+d_1 & & \\ \lambda_2\lambda_0 & \lambda_2\lambda_1 & \lambda_2^2+d_2 & \\ \lambda_3\lambda_0 & \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+d_3\end{bmatrix}\right) \tag{16}$$

Again, both the Cholesky structure in Equation 15 and the factor analytic structure in Equation 16 are able to provide the same information as the unrestricted structure shown in Equation 13, in the sense that they calculate variances and covariances for all random effects. Given that the model has more than 3 random effects, the factor analytic structure also has fewer parameters (8) compared to the unrestricted or Cholesky structure (both have 10).

The Cholesky structure was semidefinite positive but ultimately inadmissible because the estimate for l_{33} was exactly 0, resulting in an estimated SES slope variance of 0. However, the factor analytic structure converged and admissible estimates were obtained for all 8 covariance parameters with no issues. The results of the factor analytic structure are compared to the each of the simplified unrestricted structures in Table 6. We converted the factor analytic parameter estimates to variances and correlations to facilitate comparisons across the models. The results in Table 6 show that the contextual effect for Minority Status is significant ($\gamma_{04}-\gamma_{20}=1.32, t(6608)=2.16, p=.031$).

This demonstrates some of the utility of the factor analytic structure, especially with many random effects. The traditional Cholesky approach was semipositive definite but had an inadmissible slope variance estimate for SES coefficient and the simplification approach could arrive at different conclusions depending on how the model were simplified without a reliable sense of which model should be preferred. Worse yet, there is rarely a guiding theory for importance of the different random slopes in cross-sectional data (unlike longitudinal data where

there is a clear hierarchy for which random effects to remove first). The factor analytic structure was able to estimate the between-school variance for all relevant predictors without issue.

Discussion

Maximal random effect structures are valuable theoretically but difficult to estimate successfully in practice, particularly when the number of observations for each independent sampling unit is limited (e.g., few repeated measures per person). As we argue in this paper, minor changes in how researchers parameterize a maximal random effect structure can have substantial impact on the ability of the estimation routine to converge to a solution. Factor analytic, Cholesky, and unrestricted structures all can be used to estimate a maximal random effect structure; however, our empirical examples and simulation showed that the factor analytic structure facilitates convergence while also improving the quality of the estimates whereas the unrestricted structure tends to struggle to overcome nonpositive definiteness. The theoretical underpinnings guiding selection of the factor analytic structure are that (1) it reduces some of the scale magnitude differences between variance-covariance parameters, (2) the loading and uniqueness parameters are less exposed to boundaries, and (3) it reduces the dimensionality of the covariance matrix when there are more than 3 random effects. We recommend that researchers consider employing a factor analytic structure for the random effect covariance matrix prior to attempting to simplify the covariance matrix if convergence may be an issue. Retaining superfluous random effects rarely causes harm (Verbeke & Molenberghs, 2000, p. 127) but including too few random effects can result in biased estimates for the remaining random effect variance-covariance parameters, in turn biasing fixed effect standard errors and leading to potentially faulty conclusions (Agresti et al., 2004; Laird & Ware, 1982).

One somewhat unexpected finding from the simulation was the discrepancy between Cholesky and factor analytic structures. Cholesky approaches are commonly employed in software and applications and enjoy a wider literature while factor analytic structures remain less celebrated. The benefits of the Cholesky structure were evident in the simulation; however, the factor analytic structure had a non-trivial improvement in convergence and also reduced the bias of random effect standard deviations in nearly all the conditions included in our simulation. This suggests that there may be new avenues for research to explore if this finding is supported more generally and if the factor analytic structure may be researcher's best chance to model a random effect covariance matrix when there are many random effects present.

In terms of software implementation, the factor analytic structure is practical for users of some software programs as it is featured as a preprogrammed covariance structure option in both SAS Proc Mixed or Proc Glimmix [TYPE = FA(1)] and in SPSS MIXED (Covariance Type = Factor Analytic: First Order, Heterogeneous). The option is not provided in the standard lme4 package for mixed effect modeling in R and is not an option in Stata's built-in mixed effect model commands. Although the factor analytic structure does not directly appear in the lme4 R package for mixed effect models, the `nlme` package provides the option to use different versions of Cholesky decompositions, with a log-Cholesky parameterization being the default. M*plus* does not provide preprogrammed covariance structures, but the factor analytic structure could be specified rather painlessly by embedding a constraints consistent with a factor analytic decomposition (similar to the bottom panel of Figure 3) via a MODEL CONSTRAINT statement.

We discussed the version of the factor analytic structure that decomposes the random effect covariance matrix using a single common factor. However, researchers are not limited to

only a single factor and can specify more factors if desired or they think that the complexity of the covariance structure cannot be accommodated by a single factor. In SAS, the TYPE = FA (*q*) statement that controls the covariance structure requests that *q* factors be used to decompose the random effect covariance matrix. We specified *q* = 1 in our examples, but larger values can be selected where $q \le r$. Underspecifying *q* could lead to bias, but this bias is likely smaller than the bias that would be incurred by omitting random effects entirely and changing the dimension of the covariance matrix. If specifying a value for *q* is a concern, one could use model comparisons to see if more factors are needed.

Additionally, researchers can omit the uniquenesses from the factor analytic structure and rely solely on the loadings to parameterize the covariance matrix. Omitting the uniquenesses is typically referred to as an FA0 structure, a benefit of which is that the result is ensured to be semipositive definite because the square of a loading cannot be below 0 (SAS Institute Inc., 2018b, p. 6610). The FA0 structure with the number of factors equal to the number of random effects (i.e., *q* = *r*) produces a Cholesky decomposition (SAS Institute Inc., 2018b, p. 6591). A potential downside of the FA0 structure is that its reliance on only the loadings to pattern the covariance matrix is somewhat implausible for *q* < *r*, which might lead to bias.

We focused on potential impacts to fixed effect standard errors, random effect covariance parameter estimates, and convergence. However, other aspects of mixed effect models are impacted by misspecification of the random effect covariance matrix. For instance, empirical Bayes predictions of random effects that are used to form person-specific trajectories include the estimated random effect covariance matrix in the calculation such that $\hat{\mathbf{u}}_i = \hat{\mathbf{G}}\mathbf{Z}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right)$. To the extent that the random effect covariance matrix is misspecified, $\hat{\mathbf{G}}$ will be incorrect and could lead to inaccurate predictions for $\hat{\mathbf{u}}_i$. Of course, if a random effect were removed to

appease convergence issues, then no empirical Bayes predications could be calculated for that effect. The extent to which alternative random effect covariance matrix parameterization could improve empirical Bayes predictions could be an additional avenue of future studies.

Another future direction may be application to growth mixture models where convergence is a more prevalent and daunting problem (Bauer & Curran, 2003; Hipp & Bauer, 2006). Infurna and Jayawickreme (2019) and Infurna and Luthar (2016) note that atheoretical simplification of the random effect covariance structure is rampant in empirical growth mixture modeling studies even though doing so adversely affects class enumeration and class assignment. McNeish and Harring (2020) have suggested marginal models as a way to improve convergence in growth mixture models, though a caveat of this approach is that researchers lose the ability to model person-specific curves. Applying a factor analytic covariance structure has been suggested for classification methods in general (McLachlan, 2011) and extending the idea to the random effect covariance matrix may present a possible way to improve convergence in growth mixture models while retaining the ability to obtain person-specific curves.

Limitations

First, the contexts for which mixed effect models are appropriate are vast and our study focused on a narrow segment of situations for which such models are appropriate. The results from our simulation were rather decisive in favoring the factor analytic structure over the unrestricted structure; however, this is not to say that these results would generalize broadly across contexts or even to other conditions within the contexts we studied. Aspects like individualized timing of measurement occasions, attrition, and varying numbers of repeated measures are common in longitudinal data but were not studied and could impact performance. Additionally, we provided a single empirical example to show that the same pattern of

convergence extends to cross-sectionally clustered data, but a full simulation would be required to quantify the degree to which a factor analytic structure may improve convergence and performance for data characteristics commonly found in cross-sectional data. This avenue could prove more beneficial as decisions about which random effects to remove in cross-sectionally clustered data are less straightforward than in longitudinal data and the ability to retain more theoretically interesting random effects could be useful for these models.

Second, our simulation generated data such that the variances of the random effects were non-zero. In practice, models with a maximal random effect structure may yield nonpositive definite covariance matrices because one or more of the coefficients simply may not vary across people or predictors completely explain any variability that exists. Our results suggest that a factor analytic structure can improve stability in estimating covariance matrices that are positive definite in the population. However, when the population covariance matrix is nonpositive definite because some random effects truly have zero variance, changing the specification will not improve the probability that an estimation routine converges and researchers should not discount the possibility that some random effects simply might not have any variability. As a diagnostic, Bates et al. (2015) suggest fitting a principal components analysis using the random effect covariance as input to determine how many components can account 100% of the variance and to reduce the number of random effects accordingly. In R, the rePCA function in the lme4 package implements this method.

Third, we focused on models with continuous outcomes but mixed effect models are also popular when outcomes are discrete and convergence presents a greater challenge in such contexts because estimation is far more difficult because the random effects do not integrate out of the likelihood (McNeish, 2016; Schoeneberger, 2016). Furthermore, because the random

effects cannot be integrated out of the likelihood, fixed effect estimates are susceptible to bias when there are misspecifications in the random effect structure (e.g., Bauer & Sterba, 2011; Blozis & Harring, 2018; Harring & Liu, 2016). A factor analytic structure can be similarly applied in mixed effects models for binary or count outcomes and is included as a preprogrammed option in mainstream software like SAS Proc Glimmix. Exploring whether a factor analytic structure can similarly improve convergence in models with discrete outcomes could serve as an interesting follow up to the current study.

Fourth, we only considered restricted maximum likelihood estimation whereas full maximum likelihood estimation is another option that can be employed and is popular for models fit in structural equation modeling software. Restricted and full maximum likelihood are equivalent at larger sample sizes but will diverge at smaller sample sizes (McNeish, 2017). With smaller sample sizes, full maximum likelihood tends to provide downwardly biased estimates of variance components and fixed effect standard errors (Maas & Hox, 2005). From these properties, our speculation would be that nonpositive definiteness with full maximum likelihood would be encountered more often than with restricted maximum likelihood because downwardly biased variance component estimates would be more likely to be near 0.

Concluding Remarks

Retaining many random effects is theoretically desirable and allowing each of these random effects to covary is a recommended practice that increases the quality of estimated quantities. As many empirical researchers are aware, this is easier said than done and nonconvergence can quickly derail plans to use a maximal random effect structure. How researchers specify a maximal random effect structure can have pronounced consequences on the ability of the estimation routine to converge to a solution. The typical approach of using an

unrestricted structure is susceptible to scale magnitude differences and boundaries on the parameter space. Our simulations showed that this can lead to low convergence rates. Moreover, the typical remedy of simplifying the covariance structure can lead to poor confidence interval coverage rates for fixed effects (and an expected increase in the Type-I error rate for null effects). Instead, alternative parameterizations can help to mitigate some of these estimation issues. Ultimately, we found that a factor analytic structure best improves both convergence and performance.

References

Agresti, A., Caffo, B., & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis*, *47*(3), 639–653. https://doi.org/10.1016/j.csda.2003.12.009

Baird, R., & Maxwell, S. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods*, *21*(2), 175–188. http://dx.doi.org /10.1037/met0000070

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*(4), 1281–1311. https://www.jstor.org/stable/24306780

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338–363. https://doi.org/10.1037/1082-989X.8.3.338

Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: performance of alternative specifications and methods of estimation. *Psychological Methods*, *16*(4), 373–390. https://doi.org/10.1037/a0025813

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*(1), 30–52. https://doi.org/10.1037/1082-989X.9.1.30

Blozis, S. A. (2004). Structured latent curve models for the study of change in multivariate repeated measures. *Psychological Methods*, *9*(3), 334–353. https://doi.org/10.1037/1082-989X.9.3.334

Blozis, S. A., & Harring, J. R. (2018). Fitting nonlinear mixed-effects models with alternative residual covariance structures. *Sociological Methods & Research*, advanced online publication. https://doi.org/10.1177/0049124118789718

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York: Guilford.

Cheng, J., Edwards, L. J., Maldonado-Molina, M. M., Komro, K. A., & Muller, K. E. (2010). Real longitudinal data analysis for real people: Building a good enough mixed model. *Statistics in Medicine*, *29*(4), 504–520. https://doi.org/10.1002/sim.3775

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, *31*(3), 371–403. https://doi.org/10.1207/s15327906mbr3103_6

Cudeck, R., & du Toit, S. H. C. (2002). A version of quadratic regression with interpretable parameters. *Multivariate Behavioral Research*, *37*(4), 501–519. https://doi.org/10.1207/S15327906MBR3704_04

Cudeck, R., Klebe, K. J., & Henly, S. J. (1993). A simple Gauss-Newton procedure for covariance structure analysis with high-level computer languages. *Psychometrika*, *58*(2), 211-232.

Daniels, M. J., & Zhao, Y. D. (2003). Modelling the random effects covariance matrix in longitudinal data. *Statistics in Medicine*, *22*(10), 1631-1647. https://doi.org/10.1002/sim.1470

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. New York: Springer.

Diallo, T. M. O., Morin, A. J. S., & Parker, P. D. (2014). Statistical power of latent growth curve models to detect quadratic growth. *Behavior Research Methods*, *46*(2), 357–371. https://doi.org/10.3758/s13428-013-0395-1

Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). *Analysis of longitudinal data*. New York: Oxford University Press.

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *ArXiv:1701.04858 [Stat]*. http://arxiv.org/abs/1701.04858

Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, *65*, 457-483. https://doi.org/10.1093/biomet/65.3.457

Elzo, M. A. (1996). Unconstrained procedures for the estimation of positive definite covariance matrices using restricted maximum likelihood in multibreed populations. *Journal of Animal Science*, *74*(2), 317-328. https://doi.org/10.2527/1996.742317x

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research*, *37*(3), 379–403. https://doi.org/10.1207/S15327906MBR3703_4

Fisher, C. I., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2015). Women's preference for attractive makeup tracks changes in their salivary testosterone. *Psychological Science*, *26*, 1958-1964. https://doi.org/10.1177%2F0956797615609900

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley.

Gibbons, R. D., & Bock, R. D. (1987). Trend in correlated proportions. *Psychometrika*, *52*(1), 113–124. https://doi.org/10.1007/BF02293959

Gill, J., & King, G. (2004). What to do when your Hessian is not invertible: Alternatives to model respecification in nonlinear estimation. *Sociological Methods & Research*, *33*(1), 54-87.

Gomez, E. V., Schaalje, G. B., & Fellingham, G. W. (2005). Performance of the Kenward–Roger

method when the covariance structure is selected using AIC and BIC. *Communications in Statistics—Simulation and Computation*, *34*(2), 377-392.

Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. New York: Guilford.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*(5), 1357–1371. https://doi.org/10.1111/j.1467-8624.2011.01630.x

Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, *30*(22), 2696–2707. https://doi.org/10.1002/sim.4293

Harring, J. R., & Liu, J. (2016). A Comparison of estimation methods for nonlinear mixed-effects models under model misspecification and data sparseness: A simulation study. *Journal of Modern Applied Statistical Methods*, *15*(1), 539–569. https://doi.org/10.22237/jmasm/1462076760

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley.

Hedeker, D., & Gibbons, R. D. (1996). MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine*, *49*(3), 229-252. https://doi.org/10.1016/0169-2607(96)01723-3

Hedeker, D., & Mermelstein, R. J. (1998). A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research*, *33*(4), 427–455. https://doi.org/10.1207/s15327906mbr3304_1

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*(1), 36–53. https://doi.org/10.1037/1082-989X.11.1.36

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York: Routledge.

Hox, J., & McNeish, D. (2020). Small samples in multilevel modeling. In R. van de Schoot & M. Milocevic (Eds.), *Small Sample Size Solutions* (pp. 215–225). Routledge. https://doi.org/10.4324/9780429273872-18

Hox, J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Rahway, NJ. Routledge.

Infurna, F. J., & Jayawickreme, E. (2019). Fixing the growth illusion: New directions for research in resilience and posttraumatic growth. *Current Directions in Psychological Science*, *28*(2), 152–158. https://doi.org/10.1177/0963721419827017

Infurna, F. J., & Luthar, S. S. (2016). Resilience to major life stressors is not as common as thought. *Perspectives on Psychological Science*, *11*(2), 175–194. https://doi.org/10.1177/1745691615621271

Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, *42*(4), 805–820. https://doi.org/10.2307/2530695

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. https://doi.org/10.2307/2533558

Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics: Simulation and Computation, 27* (3), 591-604.

Kiernan, K. (2018). Insights into using the GLIMMIX procedure to model categorical outcomes with random effects. *SAS Institute*, SAS2179-2018.

Kohli, N., Peralta, Y., & Bose, M. (2019). Piecewise random-effects modeling software programs. *Structural Equation Modeling*, *26*(1), 156–164. https://doi.org/10.1080/10705511.2018.1516507

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974. JSTOR. https://doi.org/10.2307/2529876

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lindstrom, M. J., & Bates, D. M. (1988). Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, *83*(404), 1014-1022. https://doi.org/10.1080/01621459.1988.10478693

Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse Wishart and separation-strategy priors for Bayesian estimation of covariance parameter matrix in growth curve analysis. *Structural Equation Modeling*, *23*(3), 354–367. https://doi.org/10.1080/10705511.2015.1057285

Lu, K., & Mehrotra, D. V. (2010). Specification of covariance structure in longitudinal data analysis for randomized clinical trials. *Statistics in Medicine*, *29*, 474-488.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92.

McLachlan, G. J. (2011). Commentary on Steinley and Brusco (2011): Recommendations and cautions. *Psychological Methods*, *16*(1), 80–81. https://doi.org/10.1037/a0021141

McNeish, D. (2016). Estimation methods for mixed logistic models with few clusters. *Multivariate Behavioral Research*, *51*, 790-804. https://doi.org/10.1080/00273171.2016.1236237

McNeish, D. (2017). Small sample methods for multilevel modeling: a colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research*, *52*(5), 661–670. https://doi.org/10.1080/00273171.2017.1344538

McNeish, D., Dumas, D. G., & Grimm, K. J. (2019). Estimating new quantities from longitudinal test scores to improve forecasts of future performance. *Multivariate Behavioral Research*, advance online publication. https://doi.org/10.1080/00273171.2019.1691484

McNeish, D., & Harring, J. (2019). Covariance pattern mixture models: Eliminating random effects to improve convergence and performance. *Behavior Research Methods*, *52*, 947-979.. https://doi.org/10.3758/s13428-019-01292-4

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, *51*(4), 495–518. https://doi.org/10.1080/00273171.2016.1167008

Meyer, K. (2009). Factor-analytic models for genotype × environment type problems and structured covariance matrices. *Genetics Selection Evolution*, *41*(1), 21. https://doi.org/10.1186/1297-9686-41-21

Meyer, K., & Kirkpatrick, M. (2005). Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. *Genetics Selection Evolution*, *37*(1), 1. https://doi.org/10.1186/1297-9686-37-1-1

Miyazaki, Y., & Frank, K. A. (2006). A hierarchical linear model with factor analysis structure at level 2. *Journal of Educational and Behavioral Statistics*, *31*(2), 125–156. https://doi.org/10.3102/10769986031002125

Mulaik, S. A. (2009). *Foundations of factor analysis*. Boca Raton, FL: CRC Press.

Oman, S. D. (1991). Multiplicative effects in mixed model analysis of variance. *Biometrika*, *78*(4), 729–739. https://doi.org/10.1093/biomet/78.4.729

Pinheiro, J. C., & Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, *6*(3), 289-296. https://doi.org/10.1007/BF00140873

Piepho, H. P. (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics*, *53*(2), 761–766. https://doi.org/10.2307/2533976

Piepho, H. P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics*, *97*(1), 195–201. https://doi.org/10.1007/s001220050885

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, *86*(3), 677-690. https://doi.org/10.1093/biomet/86.3.677

Preacher, K. J., & Hancock, G. R. (2015). Meaningful aspects of change as novel random coefficients: A general method for reparameterizing longitudinal models. *Psychological Methods*, *20*(1), 84–101. https://doi.org/10.1037/met0000028

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, *52*(1), 501-525.

Rights, J. D., Preacher, K. J., & Cole, D. A. (2020). The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects. *British Journal of Mathematical and Statistical Psychology*, advance online publication.

SAS Institute Inc. (2018a). The GLIMMIX procedure. In *SAS/STAT® 15.1 User's Guide* (pp. 3616–3953). Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2018b). The MIXED Procedure. In *SAS/STAT® 15.1 User's Guide* (pp. 6536–6721). Cary, NC: SAS Institute Inc.

Schoeneberger, J. A. (2016). The impact of sample size and other factors when estimating multilevel logistic models. *The Journal of Experimental Education*, *84*(2), 373–397. https://doi.org/10.1080/00220973.2015.1027805

Schuster, C., & von Eye, A. (1998). Determining the meaning of parameters in multilevel models for longitudinal data. *International Journal of Behavioral Development*, *22*, 475-491. https://doi.org/10.1080/016502598384234

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *23*, 323-355. https://doi.org/10.3102/10769986023004323

Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, *57*(4), 1138–1147. https://doi.org/10.1111/j.0006-341X.2001.01138.x

Stroup, W. W., Milliken, G. A., Claassen, E. A., & Wolfinger, R. D. (2018). *SAS for mixed models: Introduction and basic applications*. Cary, NC: SAS Institute.

Van der Elst, W., Hermans, L., Verbeke, G., Kenward, M. G., Nassiri, V., & Molenberghs, G. (2016). Unbalanced cluster sizes and rates of convergence in mixed-effects models for clustered data. *Journal of Statistical Computation and Simulation*, *86*, 2123-2139.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag. https://doi.org/10.1007/978-1-4419-0300-6

Vermunt, J. K. (2007). Growth models for categorical response variables: Standard, latent-class, and hybrid approaches. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 139-158). Mahwah, NJ: Erlbaum.

Wang, J., Wang, P., Hedeker, D., & Chen, L. (2019). A multivariate mixed-effects selection model framework for batch-processed proteomics data with non-ignorable missingness. *Biostatistics*, *20*(4), 648-665.

Wang, L., Yang, M., & Liu, X. (2019). The impact of over-simplifying the between-subject covariance structure on inferences of fixed effects in modeling nested data. *Structural Equation Modeling*, *26*(1), 1–11. https://doi.org/10.1080/10705511.2018.1489725

Weiss, R. E. (2005). *Modeling longitudinal data*. New York: Springer.

Wolfinger, R. D. (1996). Heterogeneous variance: Covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, *1*(2), 205–230. https://doi.org/10.2307/1400366

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: Sage.

Table 1

Model estimates from BPRS data with simplified covariance matrix

Effect	Notation	Estimate	SE	*p*
Fixed Effects				
Week 8 Intercept, Group 1	γ_{00}	28.30	2.53	<.01
Week 8 Intercept, Group 2	$\gamma_{00}+\gamma_{01}$	33.97	2.53	<.01
ΔWeek 8 Intercept	γ_{01}	5.68	3.58	.12
Slope at Week 8, Group 1	γ_{10}	−1.15	0.60	.05
Slope at Week 8, Group 2	$\gamma_{10}+\gamma_{11}$	1.35	0.60	.03
ΔSlope at Week 8	γ_{11}	2.50	0.84	<.01
Quadratic Slope, Group 1	γ_{20}	0.20	0.07	<.01
Quadratic Slope, Group 2	$\gamma_{20}+\gamma_{21}$	0.41	0.07	<.01
ΔQuadratic Slope	γ_{21}	0.21	0.10	.03
Covariance Structure				
Intercept Variance	τ_{00}	118.91		
Linear Slope Variance	τ_{11}	2.06		
Intercept, Linear Correlation	ρ_{10}	0.40		
Residual Variance, Week 8	σ^2	11.80		
Heteroskedasticity Effect	ω	−0.22		

Note: Treatment Group 1 was used as the reference group. The Δ parameters are calculated by (Group 2 – Group 1). The Heteroskedasticity Effect is a log-linear model. Since Week 8 is coded as 0 so time is comprised of negative values, −0.22 means that that the residual variance at time *t* is $e^{0.22}=1.25$ times the residual variance at time $t+1$. For instance, the residual variance at Week 7 would be equal to the residual variance at Week 8 times 1.25, $(11.80\times 1.25)=14.70$.

Table 2

Model estimates for BPRS data using factor analytic covariance structure

Effect	Notation	Estimate	SE	p
Fixed Effects				
Week 8 Intercept, Group 1	γ_{00}	28.28	2.59	<.01
Week 8 Intercept, Group 2	$\gamma_{00}+\gamma_{01}$	33.98	2.59	<.01
ΔWeek 8 Intercept	γ_{01}	5.70	3.67	.11
Slope at Week 8, Group 1	γ_{10}	−1.17	0.67	.09
Slope at Week 8, Group 2	$\gamma_{10}+\gamma_{11}$	1.35	0.67	.05
ΔSlope at Week 8	γ_{11}	2.52	0.94	.01
Quadratic Slope, Group 1	γ_{20}	0.19	0.09	.04
Quadratic Slope, Group 2	$\gamma_{20}+\gamma_{21}$	0.41	0.09	<.01
ΔQuadratic Slope	γ_{21}	0.22	0.13	.09
Covariance Structure				
Intercept Loading	λ_0	6.95		
Linear Slope Loading	λ_1	2.16		
Quadratic Slope Loading	λ_2	0.21		
Intercept Uniqueness	d_0	83.14		
Linear Slope Uniqueness	d_1	0.00		
Quadratic Slope Uniqueness	d_2	0.03		
Residual Variance, Week 8	σ^2	10.74		
Heteroskedasticity Effect	ω	−0.22		

Note: Treatment Group 1 was used as the reference group. The Δ parameters are calculated by (Group 2 – Group 1). The Heteroskedasticity Effect is a log-linear model. Since Week 8 is coded as 0 so time is comprised of negative values, −0.22 means that that the residual variance at time t is $e^{0.22}=1.25$ times the residual variance at time $t+1$. For instance, the residual variance at Week 7 would be equal to the residual variance at Week 8 times 1.25, $(10.74\times1.25)=13.43$.

Table 3

Percentage of converged replications by simulation condition

	N = 50				N = 100				N = 200			
RM	UNR	OP	CD	FA	UNR	OP	CD	FA	UNR	OP	CD	FA
4 (narrow)	14	44	39	44	16	44	45	53	16	45	46	59
6	25	58	58	77	31	67	67	87	37	82	82	95
4 (wide)	33	71	71	86	33	82	82	96	33	93	93	99
8	67	97	97	99	68	100	100	100	61	100	100	100

Note: N = Sample Size, RM = Repeated Measures, UNR = Unrestricted Structure with Standard Deviation-Correlation Parameterization, OP = Orthogonal Polynomial with Unrestricted Structure, CD = Cholesky Decomposition Parameterization of Unrestricted Structure, FA = Factor Analytic Structure.

Table 4

Percent relative bias for each standard deviation or correlation parameter by simulation condition and random effect covariance matrix structure

N	RM	Element	UNR	CD	FA
50	4	Int SD	0	0	0
		Linear SD	84	169	27
		Quadratic SD	324	1349	145
		Int, LS Corr	−12	−27	0
		Int, QS Corr	−22	−60	26
		LS, QS Corr	8	0	25
50	6	Int SD	0	0	0
		Linear SD	24	28	4
		Quadratic SD	50	90	14
		Int, LS Corr	−4	−7	6
		Int, QS Corr	−5	−19	11
		LS, QS Corr	5	−2	−9
50	8	Int SD	1	0	0
		Linear SD	4	1	−2
		Quadratic SD	5	2	−3
		Int, LS Corr	−1	−1	3
		Int, QS Corr	4	0	8
		LS, QS Corr	1	−2	−5

N	RM	Element	UNR	CD	FA
100	4	Int SD	−1	2	0
		Linear SD	67	121	12
		Quadratic SD	259	919	88
		Int, LS Corr	−20	−19	10
		Int, QS Corr	−40	−53	−16
		LS, QS Corr	6	−2	−29
100	6	Int SD	0	1	1
		Linear SD	15	19	2
		Quadratic SD	32	54	4
		Int, LS Corr	−4	−4	6
		Int, QS Corr	5	−10	19
		LS, QS Corr	4	0	−2
100	8	Int SD	0	−1	0
		Linear SD	4	2	0
		Quadratic SD	5	3	−1
		Int, LS Corr	2	1	3
		Int, QS Corr	5	1	8
		LS, QS Corr	1	−1	−1

N	RM	Element	UNR	CD	FA
200	4	Int SD	0	1	0
		Linear SD	47	81	4
		Quadratic SD	191	619	50
		Int, LS Corr	−14	−17	11
		Int, QS Corr	−33	−52	0
		LS, QS Corr	4	−4	−23
200	6	Int SD	0	0	0
		Linear SD	8	7	−1
		Quadratic SD	19	22	−4
		Int, LS Corr	−3	−1	5
		Int, QS Corr	−4	−1	18
		LS, QS Corr	1	0	−1
200	8	Int SD	0	0	0
		Linear SD	2	1	0
		Quadratic SD	2	1	−1
		Int, LS Corr	1	0	2
		Int, QS Corr	3	0	5
		LS, QS Corr	1	0	−1

Note: N = Sample Size, RM = Repeated Measures, UNR = Unrestricted Structure with Standard Deviation-Correlation Parameterization, CD = Cholesky Decomposition, FA = Factor Analytic Structure, Int = Intercept, LS = Linear Slopes, QS = Quadratic Slope, SD = Standard Deviation, Corr = Correlation. Relative bias is only calculated for replications that converged for each method, so values across methods are not based on the same number of replications.

Table 5

Fixed effect 95% confidence interval coverage rates by simulation conditions and random effect covariance matrix structure

N	RM	Effect	UNR	UNR2	CD	FA	N	RM	Effect	UNR	UNR2	CD	FA	N	RM	Effect	UNR	UNR2	CD	FA
50	4	Time	97	94	91	94	100	4	Time	94	95	96	96	200	4	Time	95	94	96	94
		$Time^2$	**98**	93	92	94			$Time^2$	96	95	96	96			$Time^2$	95	95	97	95
		$Time^2$×Treat	97	95	97	97			$Time^2$×Treat	94	96	96	95			$Time^2$×Treat	93	95	95	95
		Time×Treat	96	95	97	96			Time×Treat	**98**	93	97	96			Time×Treat	96	93	96	95
		Treat	94	94	96	96			Treat	95	94	95	95			Treat	97	95	96	96
50	6	Time	96	**92**	95	94	100	6	Time	95	**92**	95	95	200	6	Time	97	93	96	96
		$Time^2$	97	94	96	96			$Time^2$	96	**92**	95	95			$Time^2$	95	**92**	94	94
		$Time^2$×Treat	**98**	94	97	96			$Time^2$×Treat	97	93	96	95			$Time^2$×Treat	95	**92**	94	94
		Time×Treat	97	92	96	94			Time×Treat	96	**92**	96	95			Time×Treat	96	**91**	95	94
		Treat	94	94	95	95			Treat	96	94	96	96			Treat	94	94	95	95
50	8	Time	96	93	95	95	100	8	Time	95	**88**	94	94	200	8	Time	96	**89**	95	95
		$Time^2$	96	**90**	96	96			$Time^2$	95	**87**	95	95			$Time^2$	95	**85**	94	94
		$Time^2$×Treat	96	**91**	96	96			$Time^2$×Treat	96	**87**	95	95			$Time^2$×Treat	96	**90**	96	95
		Time×Treat	96	**90**	96	95			Time×Treat	96	**89**	95	95			Time×Treat	95	**91**	96	96
		Treat	94	94	95	95			Treat	96	**85**	96	96			Treat	95	96	95	95

Note: N = Sample Size, RM = Repeated Measures, UNR = Unrestricted Structure with Standard Deviation-Correlation Parameterization, UNR2 = Unrestricted 2×2 Structure With No Random Effect for the Quadratic Slope, CD = Cholesky Decomposition, FA = Factor Structure. UNR, CD, and FA are based on replications that converged in those conditions. UN2 is based on replications in which UNR did not converge. Bolded entries indicate coverage rates that are outside of reasonable limits suggested by Bradley (1978)

Table 6
Estimates for different models for the High School Beyond data

	Unrestricted: No Female or SES Random Effects			*Unrestricted: No SES Random Effect*			*Factor Analytic*		
Parameter	Estimate	SE	*p*	Estimate	SE	*p*	Estimate	SE	*p*
Fixed Effects									
Intercept	11.70	.20	<.01	11.70	.20	<.01	11.71	.20	<.01
Female	−1.07	.17	<.01	−1.08	.19	<.01	−1.08	.19	<.01
Minority	−3.82	.34	<.01	−3.83	.34	<.01	−3.84	.34	<.01
SES	2.26	.15	<.01	2.26	.15	<.01	2.26	.15	<.01
School Mean Female	−2.06	.50	<.01	−2.06	.50	<.01	−2.08	.50	<.01
School Mean Minority	−2.65	.51	<.01	−2.62	.51	<.01	−2.52	.51	<.01
School Mean SES	3.92	.39	<.01	3.94	.39	<.01	3.93	.39	<.01
Private	2.11	.32	<.01	2.10	.32	<.01	2.10	.32	<.01
Size	0.07	.02	<.01	0.06	.02	<.01	0.06	.02	<.01
Female × Size	−0.07	.03	<.01	−0.07	.03	.01	−0.07	.03	.01
Minority × Private	2.06	.49	<.01	2.09	.48	<.01	2.10	.48	<.01
SES × Private	−0.95	.22	<.01	−0.95	.22	<.01	−0.95	.23	<.01
Female Contextual	−1.00	.53	.06	−0.98	.53	.07	−1.00	.53	.06
Minority Contextual	1.17	.61	.06	1.21	.61	.05	1.32	.61	.03
SES Contextual	1.67	.42	<.01	1.68	.42	<.01	1.67	.42	<.01
Covariance Structure									
Intercept Variance	1.73			1.74			1.74		
Female Slope Variance	---			0.59			0.59		
Minority Slope Variance	0.83			0.81			0.81		
SES Slope Variance	---			---			0.08		
Int, Female Correlation	---			−.19			−.25		
Int, Minority Correlation	.32			.31			.34		
Int, SES Correlation	---			---			.12		
Female, Minority Correlation	---			−.26			−.19		
Female, SES Correlation	---			---			−.07		
Minority, SES Correlation	---			---			.09		
Residual Variance	35.42			35.32			35.29		

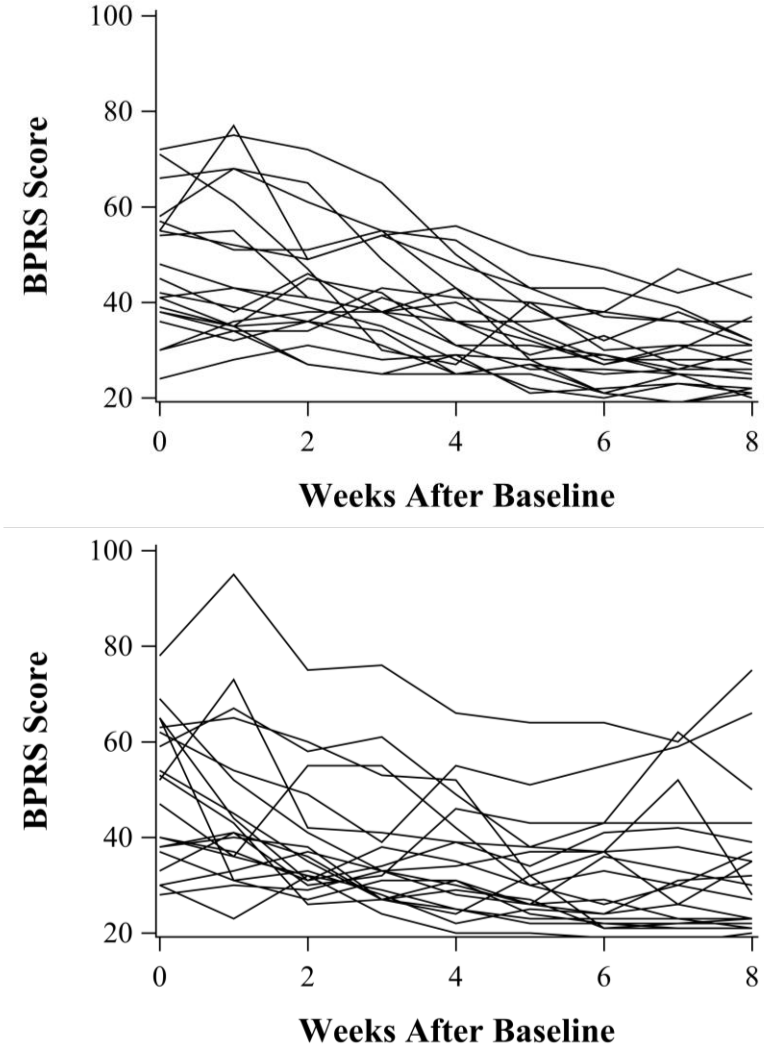

Figure 1. Empirical trajectory plots for Treatment Group 1 (top) and Treatment Group 2 (bottom)

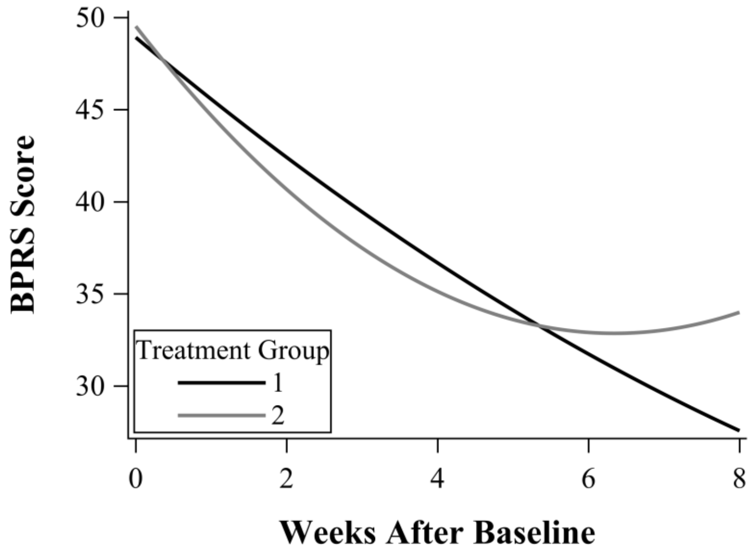

Figure 2. Average fitted trajectories for each treatment group

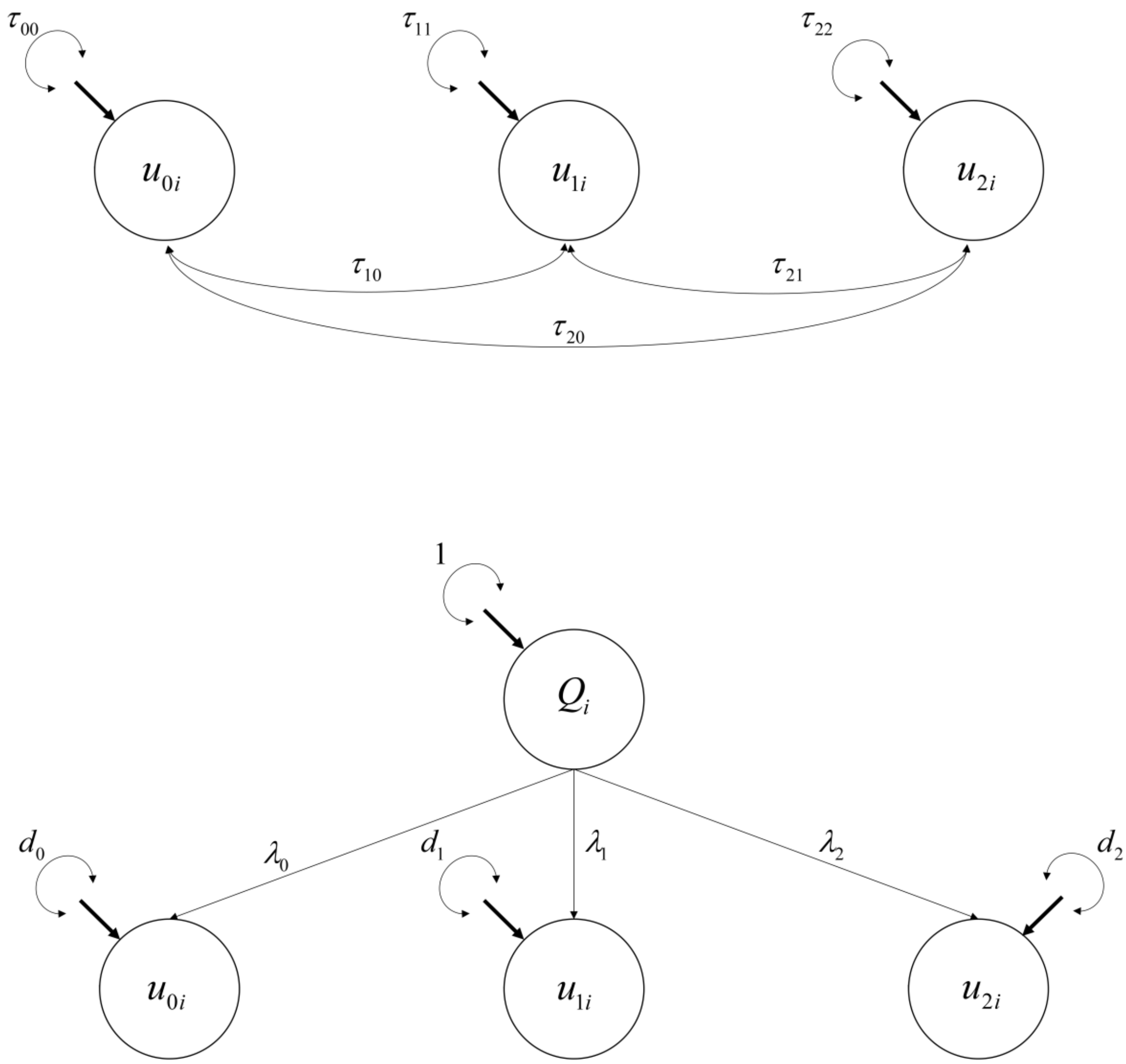

Figure 3. Comparison of path diagrams for random effect covariance matrix specifications for an unrestricted structure (top) and a factor analytic structure (bottom)